Elucidating Complex Phenotypes and the Complex Regulatory Networks That Create Phenotypes: Progress in Several Analysis Approaches

Zhang, Bing*, Kirov, Stefan, Crawford, Oakley, Williams, Robert, Langston, Michael, Leuze, Michael, Jones, Brynn, Larimer, Frank, Goldowitz, Dan, Snoddy, Jay, Tennessee Mouse Genome Consortium, Integrative Neuroscience Initiative on Alcoholism, The Program for Comparative and Collaborative Bioinformatics

Oak Ridge National Laboratory-University of Tennessee Graduate School for Genome Science and Technology, Oak Ridge, TN, USA

Fundamental Biology Requirements and Drivers: Gregory Bateson once described information *"as a difference that makes a difference.*" A mission of bioinformatics ought to be to understand the genotypic differences that help make the phenotypic differences. Phenotypes are *variable* in populations of the same species. This includes variable susceptibility to disease in humans created, in part, by genotypic variation. Secondly, phenotypes are *complex* in any individual organisms; the emergent complexity of a single individual is created out of the same genotype and protein components during development of that individual. Thirdly, phenotypes are *diverse* across different species. It is believed that the differences that create many phenotypic differences—over both developmental and evolutionary time—are often differences in gene regulatory networks and cell-cell communication networks. Computer-based approaches to comparing the properties of regulatory networks under different genetic and environmental conditions are, we feel, a fundamental component in analyzing biological data and understanding the diverse biological phenomena.

Overall Approach and Results to Date: The bioinformation systems we are building include several major features (see also abstracts of Snoddy et al and Kirov et al.). This poster will describe several approaches to initially discover, visualize, and understand the co-regulated components in gene regulatory networks. An example of a system we are developing is a web-based data mining and visualization environment, called GO Tree Machine (GOTM) (http://genereg.ornl.gov/gotm). This is used for interpreting sets of interesting genes from microarray or other high throughput experiments using Gene Ontology hierarchies. Large-scale comparisons of data sets from different conditions are facilitated by GOTM. GOTM and other tools should have broad application in functional genomic, proteomic and large scale genetic studies from which high-throughput data are continuously generated. We will discuss the application of GOTM and other approaches to the data in the WebQTL project (http://www.webqtl.org Robert Williams et al). This collaboration can help apply statistical genetics to discovery of regulatory and interconnected networks. We are attempting to build a relevance network of genes correlated to complex phenotypes. We are exploring the uses of graph theoretical approaches to regulatory network analysis (Michael Langston et al.). We will discuss work on the Tennessee Mouse Genome Consortium phenotyping project; this project uses automated statistical analysis to study altered phenotypes and genotypes among populations of mutagenized mice.